The Old School, Brewhouse Hill, Wheathampstead,
Hertfordshire AL4 8AN, UK
www.ufaw.org.uk

Animal Welfare 2019, 28: 365-380
ISSN 0962-7286
doi: 10.7120/09627286.28.3.365

Changes in the welfare of mink (Neovison vison) with date of assessment in the winter and growth periods have limited effects on the overall WelFur categorisation

AF Marsbøll†, BIF Henriksen*†, BK Hansen‡ and SH Møller†

† Department of Animal Science, Aarhus University, Blichers Allé 20, 8830 Tjele, Denmark
‡ Kopenhagen Fur, Agro Food Park 15, 8200 Aarhus N, Denmark
* Contact for correspondence: britt.henriksen@anis.au.dk

Abstract

*In this study we investigated the robustness of the WelFur welfare assessment system for farmed mink (*Neovison vison*) to date of assessment in the winter and growth assessment periods. The prevalence of occurrences of certain measurements was hypothesised to increase with date of assessment (too thin, fur-chewing and stereotypic behaviour in the winter period and injuries, diarrhoea and exploratory mink in the growth period). The welfare was assessed on eight Danish mink farms according to the WelFur-Mink protocol. Each farm was assessed once in the nursing period (to be able to calculate WelFur-Mink scores), four times in the growth period and three times in the winter period. WelFur scores were calculated based on the assessments in the three periods: one calculation for each assessment in the winter and growth periods. The odds of fur-chewing increased with date of assessment in the winter period, and the odds of injuries, diarrhoea and exploratory mink increased with date of assessment in the growth period. The odds of too thin mink in the winter period decreased, ie the change was in the opposite direction to what was expected. The effect of these changes on the aggregated WelFur scores on the higher levels was limited, but could potentially lead to changes in the overall welfare categorisation of farms if the principle scores were close to a threshold between two categories. A potential way to eliminate the effect of date of assessment could be to develop a correction factor for the measurements that can be expected to change within each assessment period.*

Keywords: *animal welfare, feasibility, mink production, reliability, welfare assessment, WelFur*

Introduction

In any animal welfare assessment system, robustness to external factors, such as time of day and weather conditions, is important in order to ensure that the assessment is reliable and feasible. In seasonal production systems, such as mink production, animals at different stages of production (eg lactation stage or age) cannot be separated from the date of the assessment. As different stages of production may be related to different welfare risk factors, date of assessment may affect the outcome of the assessment. In this study, the robustness of the WelFur-Mink welfare assessment system for farmed mink (*Neovison vison*) to date of assessment is investigated.

WelFur-Mink was developed based on the concept of the EU project Welfare Quality® (Mononen *et al* 2012). An equivalent system was developed simultaneously for farmed foxes (blue fox [*Vulpes lagopus*] and silver fox [*V vulpes*] and their crossbreeds). In WelFur-Mink, the welfare is assessed at farm level based on a range of measurements taken on the farm (Møller *et al* 2015). All measurements are relevant to some aspect of the 12 animal welfare criteria that constitute the four principles for good animal welfare that were defined within the Welfare Quality® project (Botreau *et al* 2007). Due to the seasonal production system, a full WelFur-Mink assessment is based on three assessments: one in each of the three main production seasons. The on-farm assessment periods are the winter period from 1 January to 20 February, the nursing period from 5 May to 1 July and the growth period from 23 September to 30 November. In practice, the beginning and end of each assessment period are affected by the onset of events, such as flush-feeding (short period of restricted feeding followed by *ad libitum* feeding) in the winter period, average date of birth and onset of weaning in the nursing period, and onset of sorting and pelting in the growth period. A representative sample of the mink on the farm in each assessment period is used for the assessment of all measurements, except a few mainly management-based measurements taken at farm level. After the three assessments, each farm is assigned one of four welfare categories. The categorisation is based on a transformation of each measurement result into a standardised score on a scale from 0–100 followed by a step-wise aggregation of the

measurement scores across the three periods into 12 criteria and four principle scores (Møller *et al* 2015).

The assessments take place at times in each season where the potential welfare problems are most likely to be observed and each assessment period is only six to eight weeks long in order to limit the possible variation in welfare within each period. However, the prevalence of welfare problems may still vary with date of assessment as shown in the nursing period for several measurements with increasing welfare problems closer to weaning (Henriksen & Møller 2015). This variation in welfare with date of assessment can affect the overall categorisation of the farms in instances where the farms' principle scores are close to a threshold value for another category (Henriksen 2015). The welfare can also be expected to show some variation with date of assessment in the winter and growth periods. During the winter, breeders are slimmed in order to facilitate reproduction. If the slimming is not managed carefully, there is a risk that the mink will experience hunger as well as become too thin. As a consequence of unfulfilled feeding motivation, the mink may start performing stereotypic behaviour (Damgaard *et al* 2004) and potentially also fur-chewing (Malmkvist *et al* 2013). In the growth period, mink are fed *ad libitum* or close to *ad libitum*, and most will become heavy or obese. Sustained over-feeding may result in an increased risk of diarrhoea (Hansen 1985). There is also an increasing risk of developing fatty liver (Hunter 1996) which, in practice, has been associated with abnormal faeces. Also, as the juvenile mink become more territorial with the approach of winter, there is an increasing risk of aggression resulting in injuries, especially among group-housed mink (Hansen *et al* 2014). Juveniles' response in a temperament test has also been found to change during the growth period with more confident and exploratory mink as pelting time in November approaches (Hansen 1996; Malmkvist & Hansen 2001). Based on the literature on welfare in mink, we do not anticipate any other WelFur-Mink measurements to vary systematically during the winter and growth assessment periods.

The aim of this study was to explore whether date of assessment in winter and growth assessment periods leads to a change in welfare that needs to be taken into consideration in the WelFur assessment of mink. Based on the previous section, we hypothesised that the prevalence of too thin mink, fur-chewing and stereotypic behaviour increases with date of assessment in the winter period and that the prevalence of injuries, diarrhoea and exploratory mink increases with date of assessment in the growth period. The potential changes at measurement level may not necessarily affect the welfare scores at measurement, criteria, principal and overall category level. This is hypothesised, firstly, because the transformation into scores is not linear. Secondly, because the step-wise aggregation of the scores may somewhat diminish the effect of changes at measurement level. Thirdly, because the overall categories are broad, and the final categorisation will change only if the principle scores are close to a threshold between two categories.

Materials and methods

Study design

Eight commercial Danish mink farms, which varied in geographical location, size, feed suppliers, housing conditions and combinations of colour types, were included in the study. Each farm was assessed four times during the growth period in 2014 and three times during the winter period in 2015, according to the WelFur-Mink protocol (Møller *et al* 2015). The farms were also assessed once in the nursing period in 2014 to provide data for the calculation of WelFur scores across three periods for each farm. Data collection at farm level was only conducted at one of the assessments in each period, while data collection at animal/cage level was conducted at all assessments. For each farm and assessment period, a stratified sample was taken in order to represent the farm in relation to sex, age, colour type and housing conditions, as described in Møller *et al* (2015). The same sample was assessed repeatedly on each farm in each period. Four different assessors, two in each assessment period, were involved in the study. At each assessment, one assessor carried out the evaluation, while an assistant recorded the outcome. At the beginning of each assessment period, the two assessors calibrated by assessing one farm together. Scenarios creating doubt as regards the assessment were noted, photographed and used for continuous calibration. At some assessments, the assessors were each other's assistants, thus further increasing the calibration between assessors. Both the assessor and assistant observed stereotypic behaviour (each observed approximately half the cages in the sample). Stereotypic behaviour was observed for 2 min after habituation to the assessor either 1 h before the farms' usual (expected) time of feeding or, if this was not possible, as the last observation of the day, following the WelFur-Mink assessment protocol at the time. Feeding time and the timing of the last observation of the day varied between the farms, but observations on the same farm took place at the same time within each assessment period.

Sample size

The samples consisted of 120 cages with one mink per cage in the winter period and 90 cages with one to four mink per cage in the growth period. Cages where the number of mink changed during the periods were excluded in order to ensure that the samples consisted of the same mink for all assessments. Such changes may be caused by sick or injured mink being moved to a 'hospital section', dying or being euthanased. This may have decreased the prevalence of welfare problems in the sample but ensured that there were no changes in the sample that could interfere with the effects of time. Cages where the mink were moved to another location were also excluded in order to avoid potential effects of changes in the environment interfering with the effects of time. In the winter period, 47 cages were excluded. Only one or two cages were excluded from each farm, except two farms where seven and 32 cages were excluded, mainly because the mink were moved to a new location. The resulting sample size ranged from 88 to

https://doi.org/10.7120/09627286.28.3.365 Published online by Cambridge University Press

Table 1 The four principles of welfare with the 12 underlying welfare criteria and 22 measurements in WelFur-Mink (adjusted from Møller *et al* 2015).

Principle	Criterion	Measurement
1 Good feeding	1 Absence of prolonged hunger	1 Body condition score
	2 Absence of prolonged thirst	2 Type of watering system
		3 Functioning and cleanliness of the water-points
2 Good housing	3 Comfort around resting	4 Access to a nest-box
		5 Resting quality of the nest-box/resting area
	4 Thermal comfort	6 Protection from exceptional weather conditions
		7 Nest-box material and bedding/nesting material
	5 Ease of movement	8 Space available for moving (cage area and cage height)
3 Good health	6 Absence of injuries	9 Skin lesions or injuries to the body
	7 Absence of disease	10 Mortality
		11 Diarrhoea
		12 Lameness and impaired movement
		13 Obviously sick animals
	8 Absence of pain induced by management procedures	14 Killing methods for pelting of mink
		15 Killing methods for individual mink
4 Appropriate behaviour	9 Expression of social behaviours	16 Social housing‡
		17 Age and procedures at weaning§
	10 Expression of other behaviours	18 Stereotypic behaviour
		19 Cage enrichment
		20 Fur-chewing#
	11 Good human-animal relationship† and Positive emotional state†	21 Frequency and duration of handling and transportation
		22 Temperament test#

† The two criteria are based on the same measurements;
‡ Only assessed in the growth period;
§ Only assessed in the nursing period;
# Not assessed in the nursing period.

120 cages per farm with an average of 114 cages. In the growth period, 35 cages, one to ten cages per farm, were excluded because the number of mink in these cages changed within the period. The resulting sample size ranged from 81 to 89 cages per farm with an average of 86 cages. Missing values for individual measurements did not exclude the respective cage from the study but reduced the sample size for the actual measurement.

Variable and invariable measurements

Of the 22 measurements, nine are animal-based (which include registration of the animals' health and behaviour) and 13 are resource-based (which include registration of the animals' environment and the management of the animals). The result of the assessment of the animal-based measurements may vary within the assessment periods due to changes in the welfare within the on-farm assessment periods. The only exception is mortality which is evaluated at farm level for each season. The resource-based measurements such as 'Social housing' and 'Frequency and duration of handling and transportation' are also evaluated at farm level for the entire seasons and cannot vary within the assessment periods. As this study was limited to mink kept in the same cages throughout each assessment period, the resource-based measurements, 'Type of watering system' and 'Protection from exceptional weather conditions' and the sub-measurement, 'Nest-box insulation capacity' (a part of the measurement 'Nest-box material and bedding/nesting material'), will not vary within the assessment period. Hence, these measurements were standardised to the first assessment in each assessment period. The evaluation of the remaining resource-based measurements may vary within

Table 2 WelFur-Mink classification of farms based on the four principle scores (adjusted from Møller *et al* 2015).

Category	Required principle scores
Best current practice	Two principal scores above 80 and the remaining scores above 55
Good current practice	Two principal scores above 55 and the remaining scores above 20
Acceptable current practice	Three principal scores above 20 and the remaining score above 10
Unacceptable current practice	If the requirements for 'Acceptable current practice' is not met

Table 3 The division of the winter and growth assessment periods into sub-periods.

Assessment period	Length of assessment period	Sub-period
Winter	1 January (day 1) to 20 February (day 51) (or when flush-feeding begins)	1: day ≤ 17
		2: day ≥ 18 and ≤ 35
		3: day ≥ 36
Growth	23 September (day 1) until 30 November (day 69) (or when sorting/pelting begins	1: day ≤ 13
		2: day ≥ 14 and ≤ 27
		3: day ≥ 28 and ≤ 40
		4: day ≥ 41

the assessment periods. For example, some cages were two-storey 'climbing cages', and sometimes access to the upper storey was opened or closed, thus affecting the evaluation of the measurement, 'Space available for moving (cage area and cage height)'. Also, cage enrichments and bedding material can be added, removed or used up by the mink, thereby changing the evaluation of the measurements, 'Cage enrichment' and the sub-measurement, 'Bedding material' (a part of the measurement, 'Nest-box material and bedding/nesting material') if not replaced by the farmer. Finally, the evaluation of the sub-measurement, 'Protection from draught' (a part of the measurement, 'Nest-box material and bedding/nesting material') and the measurement, 'Functioning and cleanliness of the water-points' may vary within the assessment periods.

Scoring and aggregations

WelFur-Mink uses 22 different measurements in the evaluation of animal welfare at farm level (Table 1). Some measurements consist of several sub-measurements as, for example, the measurement 'Protection from exceptional weather conditions' which is based on an evaluation of the protection from sun and wind and the possibility of cooling. Most measurements are assessed in all periods (eg 'Obviously sick animals') while others are period-specific (eg 'Age and procedures at weaning'). The result of each measurement taken on the farm in each assessment period (eg percentage without access to a nest-box) is first transformed into a measurement score for each period on a standardised scale ranging from 0 (worst) to 100 (best) and aggregated across the three assessment periods into 22 measurement scores. The 22 measurement scores are then aggregated into 12 scores at criteria level and further into four scores at principle level (Møller *et al* 2015). Table 1 gives an overview of which measurements are aggregated to which criteria and which criteria are aggregated to which principles. All aggregations use Choquet integrals (Møller *et al* 2015). This means that when aggregating a number of scores, the lowest score is the starting point, and this score is partly compensated for by the higher scores, depending on the assigned weights. The transformation of the measurements taken on the farm in each period into a measurement score for each period, and the weights used when going from the measurement scores for each period to measurement scores across the three assessment periods to criteria scores, were derived from experts in fur animal welfare. The weights used when going from criteria scores to principle scores were derived from the weights used in Welfare Quality®. Finally, each farm is assigned one of four welfare categories ('Best current practice', 'Good current practice', 'Acceptable current practice', or 'Unacceptable current practice') based on the four principle scores, depending on the threshold values described in Table 2. These are the same as in Welfare Quality® (Møller *et al* 2015).

Data analysis

The aim of our study was to evaluate the effect of date of assessment on the resulting welfare assessment. Date of assessment is defined as 'Day in assessment period', ie assessments on 9 January are day 9 of the winter period. Observations in the winter period were initiated on 6 January (day 6) and completed on 19 February (day 50; one day before the end of the assessment period). Observations in the growth period were initiated on 22 September (day 0; one day before the start of the assessment period) and completed on 11 November (day 50). The end of the growth period varies between farms according to their practices as regards sorting before pelting. However, in practice, there will only be a few assessments after 15 November. For the analysis, the winter period was divided into three sub-periods and the growth period into four (Table 3). The difference with sub-period in the assessed welfare was evaluated individually for the winter and growth periods and investigated both at measurement level and for the calculated WelFur scores (measurement score for each period, criteria score and principle score).

The outcome variables included in the analysis at measurement level include, 'Too thin', 'Fur-chewing' and 'Stereotypic behaviour' in the winter period and, 'Injuries', 'Diarrhoea' and 'Exploratory' in the growth period (Table 4). The variables were included as binary variables as in Henriksen and Møller (2015). Regarding, 'Injuries' and 'Fur-chewing', it should be noted that the WelFur-Mink system also takes into account the severity of the injury or fur-chewing. However, due to the relatively low prevalence of

https://doi.org/10.7120/09627286.28.3.365 Published online by Cambridge University Press

Table 4 Description of the outcome variables included in the analysis at measurement level.

Variable	Description
Too thin	Body condition is assessed on a scale from 1 (very thin) to 5 (obese). Mink in category 1 and 2 in January and category 1 in February are associated with prolonged hunger (too thin). Binary variable (too thin or not)
Fur-chewing	Fur (guard hair or wool) has been chewed off. Assessed on a scale from 0 (no/very little) to 3 (extensive fur-chewing). Binary variable (≥ 1 or 0)
Stereotypic behaviour	A repetitive, invariant behaviour without any obvious function or goal. If three repetitions of the same behaviour is observed in the two-min observation period the mink is considered stereotypic. Binary variable (stereotypic or not)
Injuries	Healed or unhealed wounds and injuries. Assessed on a scale from 0 (no injuries) to 3 (major unhealed injuries). Binary variable (≥ 1 or 0)
Diarrhoea	Diarrhoea is defined as very mucous, watery, fluent or bloody manure without form or texture. Diarrhoea is registered at cage level as it is not possible to determine how many mink in a cage are affected. Binary variable (signs of diarrhoea or not)
Exploratory	The behavioural response of the mink to a wooden tongue spatula inserted into the cage. The mink's response is assessed as exploratory, fearful, aggressive or undecided. Binary variable (exploratory or not)

mink with injuries and fur-chewing in the higher categories, it was not possible to analyse each category separately. The effect of date of assessment on the measurement variables in each period was analysed using binomial mixed models with a logit-link function with sub-period as explanatory variable. Assessor was included as a fixed effect, whereas farm, cage nested in farm (due to repeated measurements on cage) and assessment number nested in farm (in order to take variance heterogeneity between assessments into account) were included as random effects. Assessor was removed from the model when the effect was insignificant.

In order to investigate how potential changes with date of assessment at measurement level in the winter and growth periods affect the measurement scores in each period, criterion scores, principle scores and the overall classification, we calculated WelFur-Mink scores for each assessment on each farm. The calculations are described in the WelFur-Mink protocol (Møller *et al* 2015). This was carried out for one assessment period at a time, using the actual values from each assessment for the measurements that we hypothesised to change. When investigating the changes in the winter period, the calculations included the actual values for each assessment for the measurements, 'Body condition score' (percentage too thin), 'Fur-chewing' (a combination of the percentage in each category) and 'Stereotypic behaviour' (percentage stereotypic) in the winter period. When investigating the changes in the growth period, the calculations included the actual value for each assessment for the measurements, 'Skin lesions or injuries to the body' (a combination of the percentage of mink with injuries in each categories), 'Diarrhoea' (percent cages with signs of diarrhoea) and 'Temperament test' (a combination of the percentage of mink categorised as 'Exploratory', 'Fearful', 'Aggressive' and 'Undecided') in the growth period. The remaining variable measurements in the winter and growth periods were set to an average over the assessments for each farm in each period, and the invariable measurements were included with the actual values for each farm in each period.

All measurements taken in the nursing period were included with the actual values for each farm. In order to investigate whether the potential changes in the WelFur-Mink scores are affected by changes in other measurements, we repeated the above calculation but included all variable measurements in the winter and growth periods, respectively, with their actual values. The distribution of the calculated scores was approximately normal in both sets of calculations. The effect of date of assessment on the calculated scores was analysed using mixed models with sub-period as explanatory variable, assessor as a fixed effect and farm as a random effect. Assessor was removed from the model when the effect was non-significant.

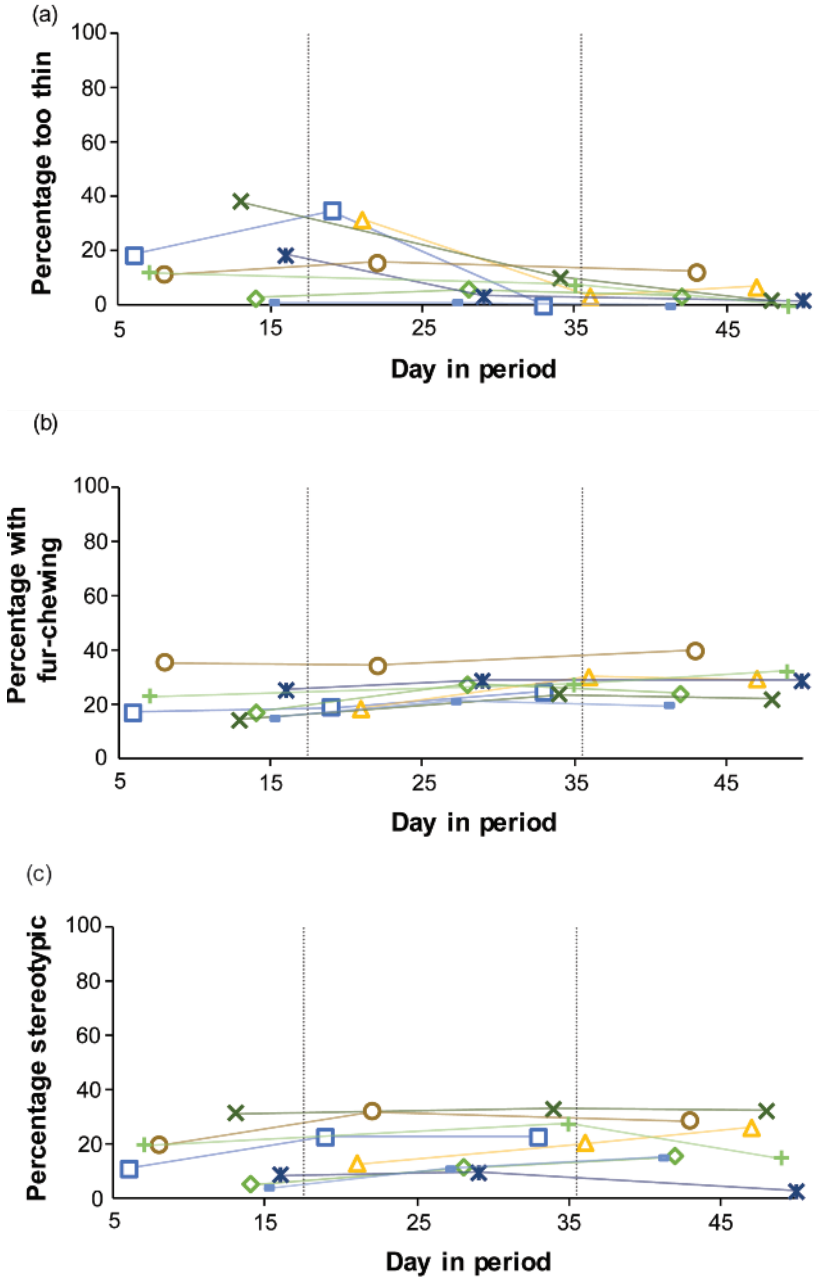
The calculations of the scores were done using the WelFur-Mink calculation tool developed by the French National Institute of Agronomic Research (INRA). The statistical analyses were performed with the package 'lme4' (Bates *et al* 2015) of the statistical software R (R Core Team 2017). Pair-wise comparisons were performed with the package 'multcomp', and the *P*-values were adjusted according to the Bonferroni method (Hothorn *et al* 2008). The limit for statistical significance was set to 5%, ie only effects where the *P*-value of the test (or the adjusted *P*-value) was less than 0.05 were declared statistically significant. All graphical presentations were made in Microsoft Excel®.

Results

Changes with date of assessment in the winter period

The prevalence of occurrences for the outcome variables that we hypothesised would change in the winter period is shown in Figure 1. Figure 1(a) indicates a decrease in the prevalence of 'Too thin' mink with assessment date, at least for some of the farms. The analysis confirmed this, as there was a significant change with sub-period in the odds of 'Too thin' mink ($\chi^2 = 31.1$, df = 2; $P < 0.001$) where the odds in sub-period 3 were 0.002 (95% CI [0.0001;0.03]) times lower than in sub-period 1 and 0.01 (95% CI [0.001;0.2])

Figure 1



Percentage of (a) too thin mink, (b) mink with fur-chewing and (c) stereotypic mink at the different assessment days in the winter period. Each symbol represents a farm and the vertical dotted lines indicates the shift between sub-periods.

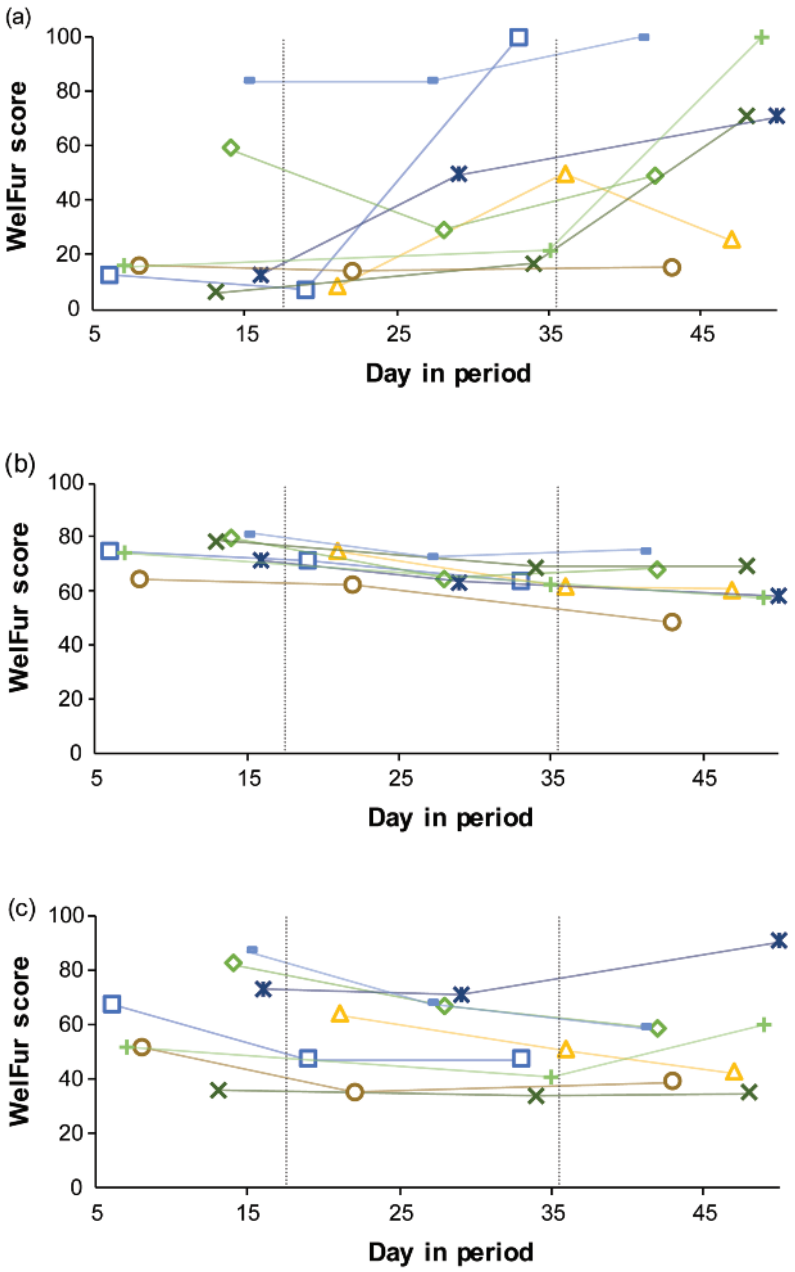
times lower than in sub-period 2. Figure 1(b) indicates a slight increase in the prevalence of 'Fur-chewing.' The analysis confirmed this, as there was a significant change with sub-period in the odds of mink with 'Fur-chewing' (χ^2 = 30.1, df = 2; $P < 0.001$) where the odds in sub-periods 2 and 3 were 4.5 (95% CI [2.3;8.8]) and 6.4 (95% CI [3.3;12.6]) times higher than in sub-period 1, respectively. Figure 1(c) indicates that the prevalence of 'Stereotypic behaviour' in general does not change with date of assessment, as some farms increase while others decrease. The analysis confirmed this, as there was no change with sub-period in the odds of 'Stereotypic behaviour' (χ^2 = 6.0, df = 2; P = 0.05).

The results from the transformation of the outcome variables that we hypothesised would change in the winter period into the measurement scores for the winter period are shown in Figure 2. Figure 2(a) shows a high variation between and within farms for 'Body condition score' in the winter period. Overall, it seems that the scores increase with assessment date. The analysis did not confirm this, as there was no change with sub-period (χ^2 = 5.8, df = 2; P = 0.06). Figure 2(b) indicates a gradual decrease in the score for 'Fur-chewing' in the winter period with date of assessment. The analysis confirmed this, as there was a significant change with sub-period (χ^2 = 21.2, df = 2;

https://doi.org/10.7120/09627286.28.3.365 Published online by Cambridge University Press

Figure 2

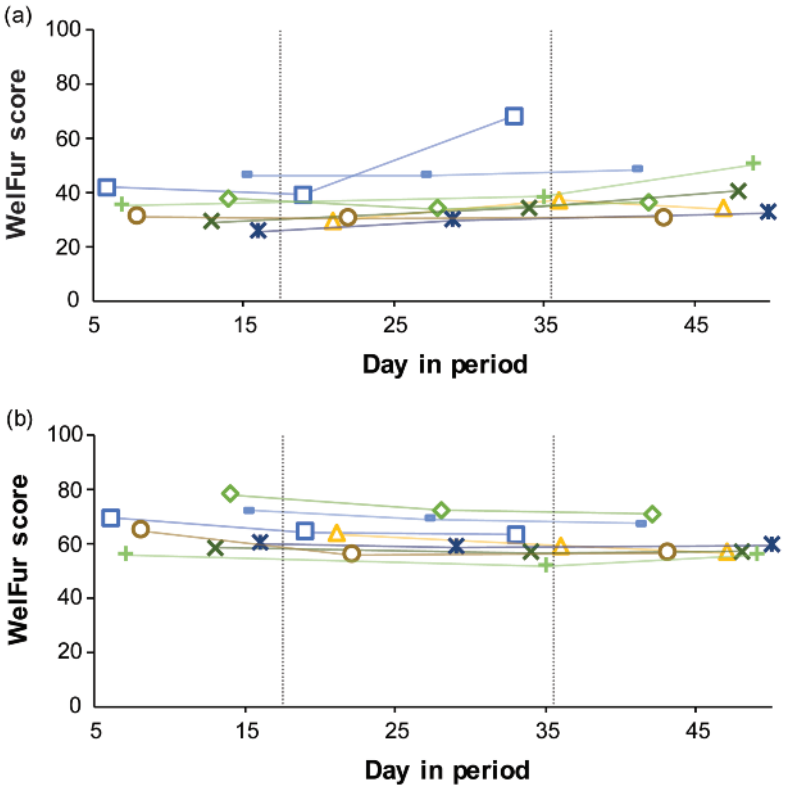
Measurement scores for (a) 'Body condition score', (b) 'Fur-chewing' and (c) 'Stereotypic behaviour' at the different assessment days in the winter period. Each symbol represents a farm and the vertical dotted lines indicate the shift between sub-periods. One hundred is the best score and 0 the worst as regards animal welfare.



$P < 0.001$) where the scores in sub-periods 2 and 3 were 8.0 points (95% CI [–12.7;–3.4]) and 13.0 points (95% CI [–18.0;–8.0]) lower than in sub-period 1, respectively. Also, the score in sub-period 3 was 5.0 points (95% CI [–9.6;–0.3]) lower than in sub-period 2. Figure 2(c) shows a high variation between and within farms for the score for 'Stereotypic behaviour' in the winter period but no general trend. The analysis showed no change in the score with sub-period ($\chi^2 = 5.6$, df = 2; $P = 0.06$).

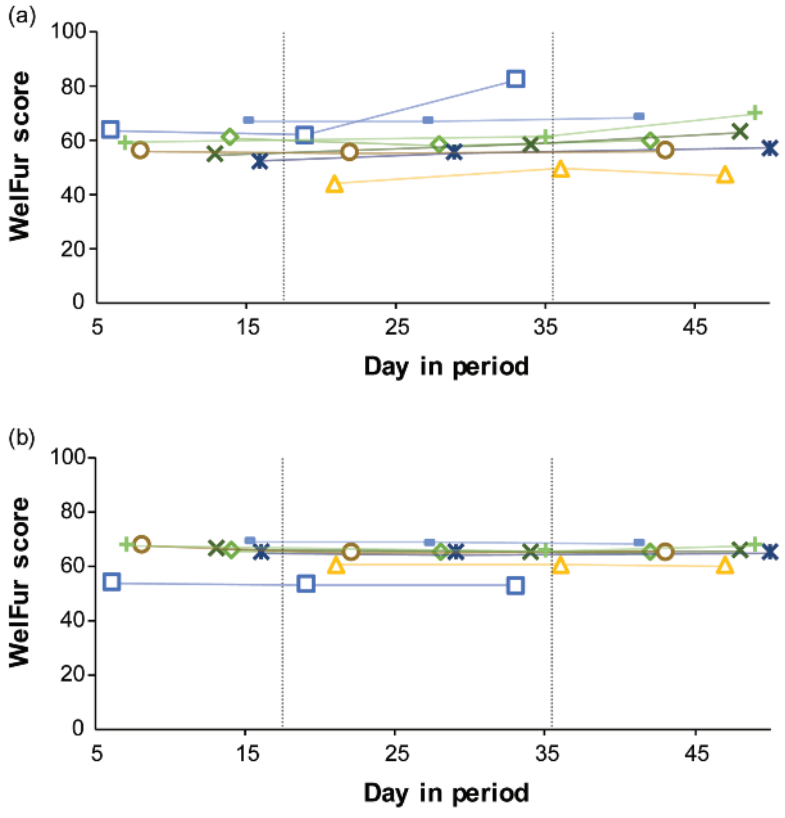
The criteria scores calculated when the measurements 'Body condition score', 'Fur-chewing' and 'Stereotypic behaviour' were included with their actual values at different assessment days in the winter period, are shown in Figure 3. Figure 3(a) indicates that there are no general or only small changes in the scores for the criterion, 'Absence of prolonged hunger' with date of assessment, except for one farm with an increase of approximately 30 points from the second to the third assessment. However, the analysis showed no change in the score with sub-period ($\chi^2 = 2.8$, df = 2; $P = 0.3$). Figure 3(b) indicates a small but general decrease in the score for the criteria 'Expression of other behaviours' with assessment date. The analysis confirmed this, as there was a significant change with sub-period ($\chi^2 = 12.5$, df = 2; $P = 0.002$) where the scores in sub-periods 2 and 3 were 4.0 points (95% CI [–6.5;-0.9]) and 4.5 points (95% CI [–7.0;–1.1]) lower, respectively, than in sub-period 1.

Figure 3



Criteria scores for (a) 'Absence of prolonged hunger' and (b) 'Expression of other behaviours' when the measurements 'Body condition score', 'Fur-chewing' and 'Stereotypic behaviour' were included with their actual values at different assessment days the in the winter period. Each symbol represents a farm and the vertical dotted lines indicates the shift between sub-periods. One hundred is the best score and 0 the worst as regards animal welfare.

Figure 4

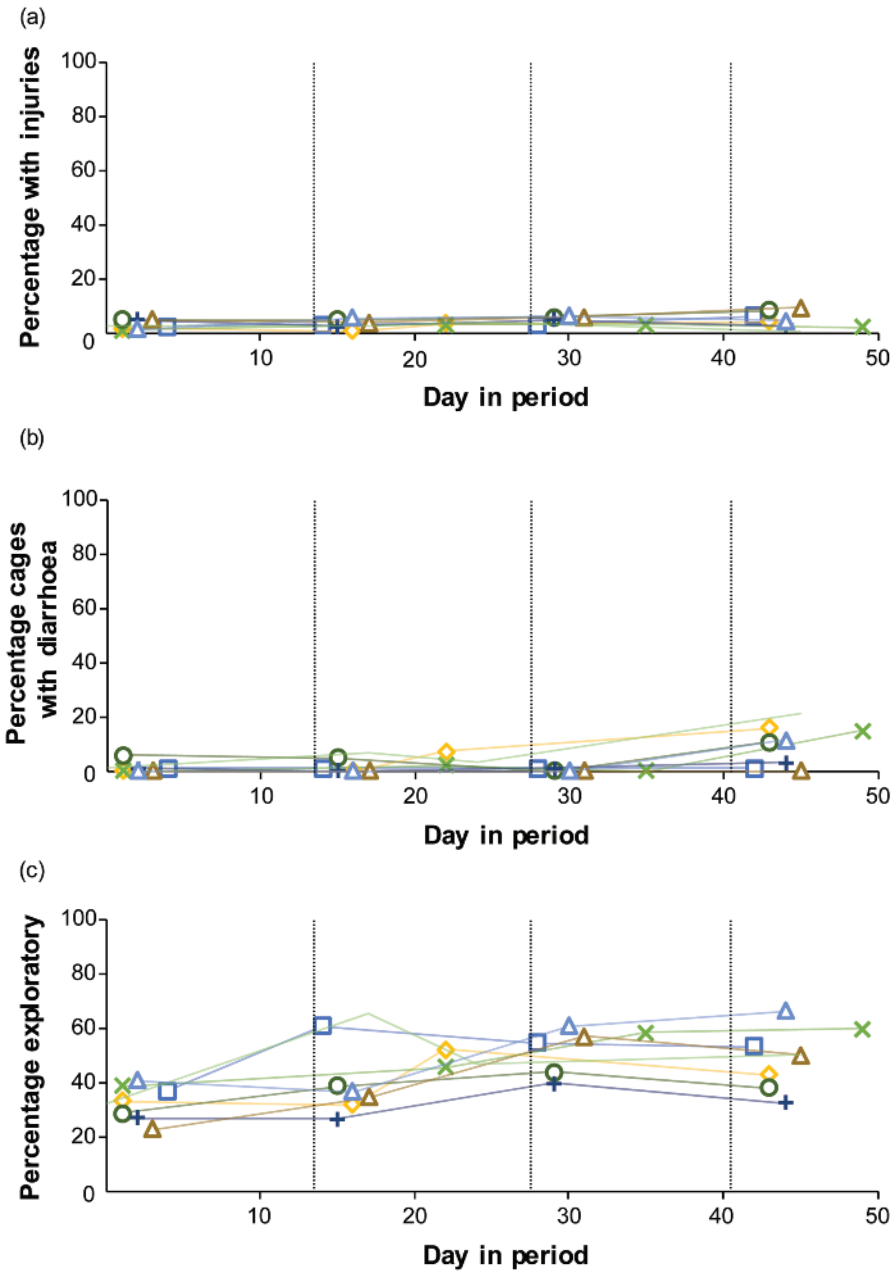


Principle scores for (a) 'Good feeding' and (b) 'Appropriate behaviour' when the measurements 'Body condition score', 'Fur-chewing' and 'Stereotypic behaviour' were included with their actual values at the different assessment days in the winter period. Each symbol represents a farm and the vertical dotted lines indicates the shift between sub-periods. One hundred is the best score and 0 the worst as regards animal welfare.

https://doi.org/10.7120/09627286.28.3.365 Published online by Cambridge University Press

Figure 5

Percentage of (a) mink with injuries, (b) cages with diarrhoea and (c) exploratory mink at the different assessment days in the growth period. Each symbol represents a farm and the vertical dotted lines indicates the shift between sub-periods.



The principle scores calculated when the measurements 'Body condition score', 'Fur-chewing' and 'Stereotypic behaviour' were included with their actual values at different assessment days in the winter period are shown in Figure 4. Figure 4(a) indicates that there are no or only small changes in the scores for the principle of 'Good feeding' with date of assessment, except for one farm where there was an increase of approximately 20 points from the second to the third assessment. The analysis showed no change with sub-period (χ^2 = 2.6, df = 2; P = 0.3). Figure 4(b) indicates that there is no change in the score for the principle of 'Appropriate behaviour' with assessment date. The analysis, however, showed that there was a change with sub-period (χ^2 = 9.0, df = 2; P = 0.01) where the score in sub-period 2 was 1.0 (95% CI [–1.6;–0.3]) lower than in sub-period 1.

The overall categorisation of the farms was not affected by the changes in the principal scores. Seven farms were classified as 'Good current practice' and one as 'Best current practice' at all assessments.

Analysing the WelFur scores that were calculated with the actual values for all variable measurements in the winter period also showed that the score for the measurement 'Cage enrichment' changed with sub-period (χ^2 = 10.6, df = 2; P = 0.005), as the score in sub-period 3 was 9.0 points (95% CI [3.0;15.0]) higher than in sub-period 2. The aggregated score for the criteria 'Expression of other behaviours' still decreased with sub-period (χ^2 = 10.6, df = 2;

$P = 0.005$), and the scores in sub-periods 2 (95% CI [–6.9;–1.1]) and 3 (95% CI [–7.1;–0.9]) were 4.0 points lower than in sub-period 1. There were no changes in any other criteria or principle scores, and the overall categorisation of the farms was the same as described above.

Changes with date of assessment in the growth period

The prevalence of occurrences for the outcome variables that we hypothesised would change in the growth period is shown in Figure 5. Figure 5(a) indicates a small increase in the prevalence of 'Injuries' with date of assessment. The analysis confirmed this, as there was a change with sub-period in the odds of mink with 'Injuries' ($\chi^2 = 9.0$, df = 3; $P = 0.03$) where the odds were 2.1 (95% CI[1.1;3.8]) times higher in sub-period 4 compared to sub-period 1. Figure 5(b) indicates that the prevalence of 'Diarrhoea' is higher at the end of the assessment period. This was confirmed by the analysis, as there was a change with sub-period in the odds of cages with 'Diarrhoea' ($\chi^2 = 17.6$, df = 3; $P < 0.001$) where the odds in sub-period 4 were 11.8 (95% CI[2.2;62.4]), 5.1 (95% CI[1.2;21.5]) and 27.9 (95% CI[2.7;291.9]) times higher than in sub-periods 1, 2 and 3, respectively. Figure 5(c) indicates a gradual increase in the prevalence of 'Exploratory' mink with date of assessment. As the prevalence of 'Exploratory' mink increased, the prevalence of both 'Fearful' and 'Undecided' mink decreased, while there were no mink categorised as 'Aggressive.' The increase in the prevalence of 'Exploratory' mink was confirmed by the analysis, as there was a change with sub-period ($\chi^2 = 31.1$, df = 3; $P < 0.001$) where the odds in sub-periods 2, 3 and 4 were 1.8 (95% CI[1.3;2.5]), 3.2 (95% CI[2.1;4.8]) and 2.2 (95% CI[1.5;3.1]) times higher, respectively, than in sub-period 1. There was also a difference between sub-periods 2 and 3, where the odds of 'Exploratory' mink in sub-period 3 were 1.8 (95% CI[1.2;2.6]) times higher than in sub-period 2.

The results from the transformation of the outcome variables that we hypothesised would change in the growth period into the measurement scores for the growth period are shown in Figure 6. Figure 6(a) indicates that there is no or only a small change with date of assessment in the score for 'Skin lesions or injuries to the body' in the growth period. The analysis showed that there was no change with sub-period in the score for 'Skin lesions or injuries to the body' in the growth period ($\chi^2 = 5.4$, df = 3; $P = 0.1$). Figure 6(b) shows a high variation between and within farms for the score for 'Diarrhoea' in the growth period. The analysis showed that there was a change with sub-period in the score for 'Diarrhoea' in the growth period ($\chi^2 = 15.4$, df = 3; $P = 0.002$) where the score in sub-period 4 was 38.6 points (95% CI [–65.7;–11.4]) lower than in sub-period 1 and 41.5 points (95% CI [–71.9;–11.2]) lower than in sub-period 3. Figure 6(c) indicates that the score for 'Temperament test' in the growth period is lower in the first part of the assessment period. This was confirmed by the analysis, as there was a change with sub-period in the score for 'Temperament test' ($\chi^2 = 43.8$, df = 3; $P < 0.001$) where the scores in sub-periods 2, 3 and 4 were 12.1 points (95% CI [7.9;16.2]), 16.4 points (95% CI [11.7;21.2]) and 14.3 points (95% CI [10.0;18.7]) higher, respectively, than in sub-period 1.

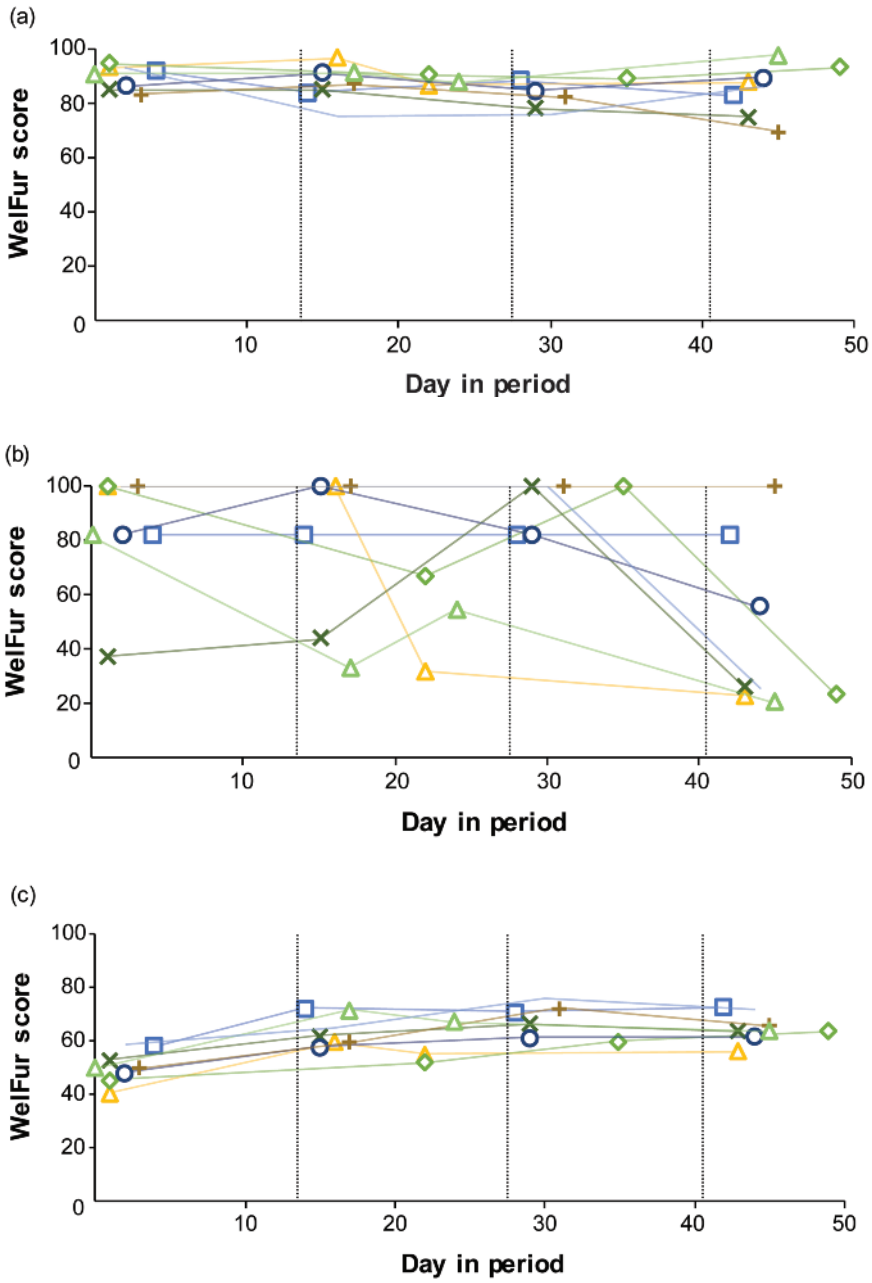
The criteria scores calculated when the measurements 'Skin lesions or injuries to the body', 'Diarrhoea' and 'Temperament test' were included with their actual values at the different assessment days in the growth period are shown in Figure 7. Figure 7(a) indicates that there is no change in the score for the criteria 'Absence of injuries' with date of assessment. This was confirmed by the analysis as there was no change with sub-period ($\chi^2 = 4.7$, df = 3; $P = 0.2$). Figure 7(b) indicates that the score for the criteria 'Absence of disease' may decrease with date of assessment. This was confirmed by the analysis as there was a change with sub-period ($\chi^2 = 17.1$, df = 3; $P < 0.001$) where the scores in sub-period 4 were 6.5 points (95% CI [–10.9;–2.0]) lower than in sub-period 1, 4.9 points (95% CI [–9.1;–0.7]) lower than in sub-period 2 and 7.8 points (95% CI [–12.6;–2.9]) lower than in sub-period 3. Figure 7(c) indicates that the score for the criteria 'Good human-animal relationship'/'Positive emotional state' is lower at the beginning of the assessment period. This was confirmed by the analysis, as there was a change with sub-period ($\chi^2 = 36.3$, df = 3; $P < 0.001$) where the scores in sub-periods 2, 3 and 4 were 6.0 points (95% CI [3.4;7.1]), 7.5 points (95% CI [4.8;9.0]) and 6.7 points (95% CI [4.2;8.1]) higher, respectively, than in sub-period 1.

The principle scores calculated when the measurements, 'Skin lesions or injuries to the body', 'Diarrhoea' and 'Temperament test' were included with their actual values at the different assessment days in the growth period are shown in Figure 8. Figure 8(a) shows a large but quite stable variation between the farms in the score for the principle 'Good health'. Even so, the analysis showed that there was a change with sub-period ($\chi^2 = 16.5$, df = 3; $P < 0.001$) where the scores in sub-period 4 were 4.7 points (95% CI [–7.8;–1.5]) lower than in sub-period 1, 3.4 points (95% CI [–6.4;–0.5]) lower than in sub-period 2 and 5.2 points (95% CI [–8.6;–1.7]) lower than in sub-period 3. Figure 8(b) indicates a slight increase with date of assessment in the score for the principle 'Appropriate behaviour'. The analysis confirmed this, as there was a change with sub-period ($\chi^2 = 36.1$, df = 3; $P < 0.001$) where the scores in sub-periods 2, 3 and 4 were 3.0 points (95% CI [1.8;4.3]), 4.1 points (95% CI [2.6;5.5]) and 3.7 points (95% CI [2.4;5.0]), respectively, higher than in sub-period 1.

For one farm, the overall categorisation was affected when the measurements, 'Skin lesions or injuries to the body', 'Diarrhoea' and 'Temperament test' were included with their actual values at the different assessment days in the growth period. The change in category was determined by changes in the principle of 'Good health' where the value was either above or below the threshold of 80 points, while all the principle scores for 'Good housing' were above 80 and 'Good feeding' and 'Appropriate behaviour' were between 55 and 80. When the actual values from the first

https://doi.org/10.7120/09627286.28.3.365 Published online by Cambridge University Press

Figure 6

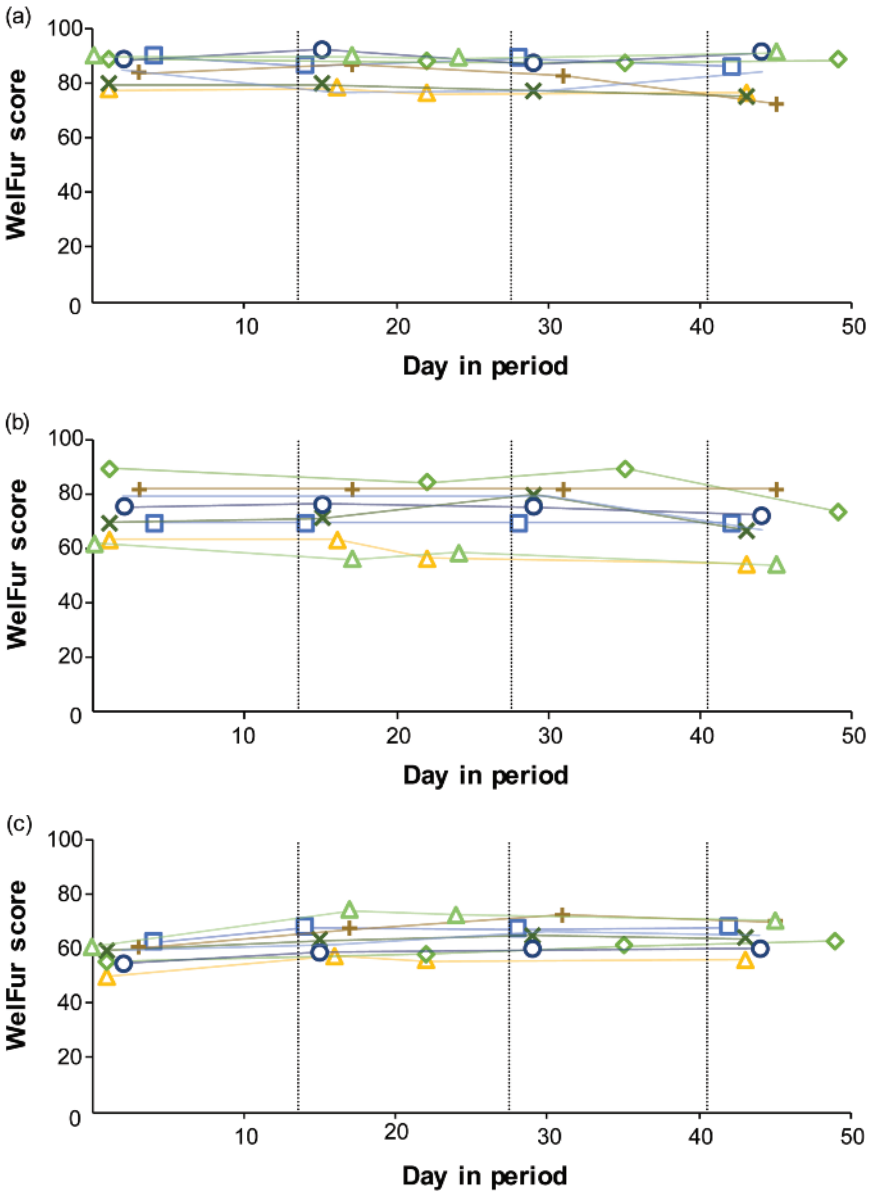


Measurement scores for (a) 'Skin lesions or injuries to the body', (b) 'Diarrhoea' and (c) 'Temperament test' at the different assessment days in the growth period. Each symbol represents a farm and the vertical dotted lines indicate the shift between sub-periods. One hundred is the best score and 0 the worst as regards animal welfare.

and fourth assessments in the growth period were used (principle scores of 78.6 and 76.1, respectively), the farm was categorised as 'Good current practice', and when the values for the second and third assessments were used (principle scores of 80.1 and 85.2, respectively), the farm was categorised as 'Best current practice'. The remaining seven farms were in the same category independent of the changes with assessment date in the growth period. Six farms were categorised as 'Good current practice' and one farm as 'Best current practice'.

Analysing the WelFur scores that were calculated with the actual values for all variable measurements in the growth period revealed a change in the measurement score for 'Nest-box material and bedding/nesting material' ($\chi^2 = 9.5$, df = 3; $P = 0.02$) where the score in sub-period 4 was 19.0 points (95% CI [2.7;35.2]) higher than in sub-period 1. Also, the criterion score for 'Thermal comfort' was found to change with sub-period ($\chi^2 = 10.3$, df = 3; $P = 0.02$) where the score in sub-period 4 was 2.8 points (95% CI [0.5;5.2]) higher than in sub-period 1. There were still changes with sub-period for the criteria 'Absence of disease' ($\chi^2 = 16.3$, df = 3; $P = 0.001$) and 'Good human-animal relationship'/'Positive emotional state' ($\chi^2 = 36.3$, df = 3; $P < 0.001$) and the principles 'Good health' ($\chi^2 = 15.6$, df = 3; $P = 0.001$) and 'Appropriate behaviour' ($\chi^2 = 34.0$, df = 3; $P < 0.001$). These changes were almost identical to the changes when only including the

Figure 7



Criteria scores for (a) 'Absence of injuries', (b) 'Absence of disease' and (c) 'Good human-animal relationship'/'Positive emotional state' when the measurements 'Skin lesions or injuries to the body', 'Diarrhoea' and 'Temperament test' were included with their actual values at the different assessment days in the growth period. Each symbol represents a farm and the vertical dotted lines indicates the shift between sub-periods. One hundred is the best score and 0 the worst as regards animal welfare.

actual values for the measurements that we expected to change. There were no changes in any other criteria or principle scores, and the overall categorisation of the farms was the same as described above.

Discussion

The results support the hypothesis that the prevalence of mink with fur-chewing increases with date of assessment in the winter period and that the prevalence of mink with injuries, cages with diarrhoea and exploratory mink increases with date of assessment in the growth period. Unexpectedly, there was no significant change in the prevalence of stereotypic behaviour in the winter period, and the prevalence of too thin mink in the winter period decreased. As expected, only the largest changes affected the WelFur scores at measurement, criteria and principle level and can potentially affect the overall categorisation of the farms.

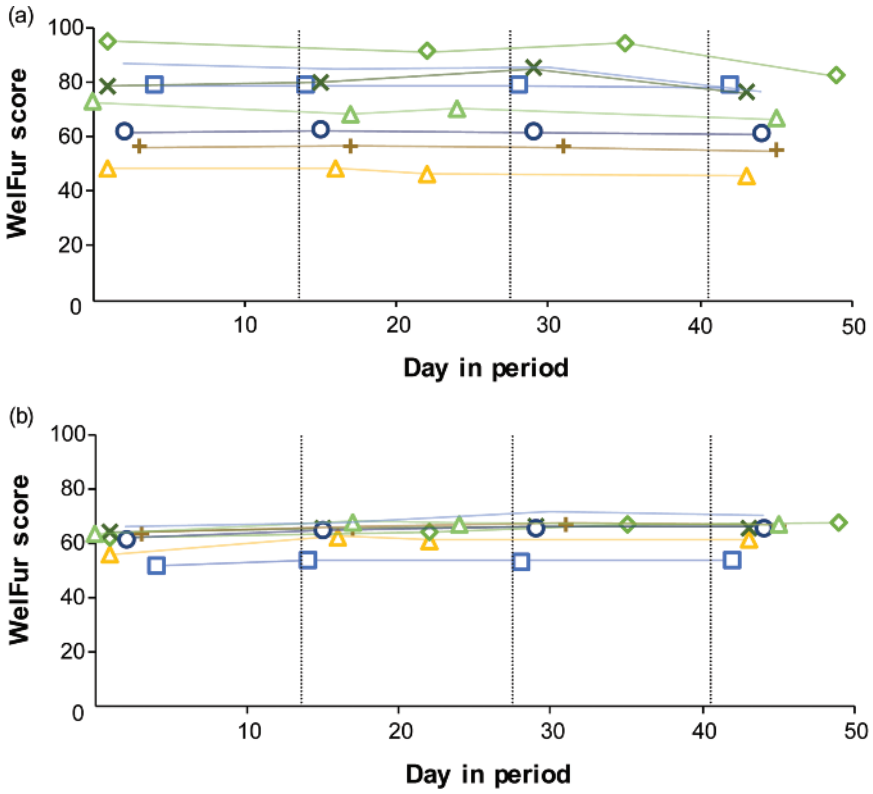
Winter period

There was a change with sub-period in the odds of too thin mink in the winter period. However, the change was in the opposite direction to what we hypothesised, as the odds in sub-period 3 were lower than in sub-period 1 and 2. The difference in the categorisation of too thin mink in January and February may be the reason for this. In January, mink in body condition 1 and 2 are considered too thin, while in February only mink in body condition 1 are considered too thin (Table 3). The reason for this categorisation is the objective of the criteria, ie to identify mink exposed to prolonged hunger. On most farms, all mink are fed *ad libitum* during the growth season in order to produce large pelts. As a result, the majority of the mink will be in body condition 4 or 5 by the end of the growth season. Based on grading of pelt quality and evaluation of body

https://doi.org/10.7120/09627286.28.3.365 Published online by Cambridge University Press

Figure 8

Principle scores for (a) 'Good health' and (b) 'Appropriate behaviour' when the measurements, 'Skin lesions or injuries to the body', 'Diarrhoea' and 'Temperament test' were included with their actual values at the different assessment days in the growth period. Each symbol represents a farm and the vertical dotted lines indicates the shift between sub-periods. One hundred is the best score and 0 the worst as regards animal welfare.



size in November, the best animals are selected as the following year's breeders. Thereafter, slimming of the selected breeding animals before mating can begin. If slimming is carried out too drastically, it is associated with prolonged hunger in WelFur-Mink. Hence, in WelFur-Mink, mink slimmed to a body condition 2 or less in January are associated with prolonged hunger. In February, there has been a longer period to slim down the mink, and only mink slimmed to a body condition 1 are associated with prolonged hunger in WelFur-Mink (Møller *et al* 2015). This means that even if the odds of mink in body condition 1 and 2 increase throughout the winter assessment period, the odds of too thin mink, according to the WelFur-Mink protocol, may decrease, as only body condition 1 is associated with prolonged hunger in February. A more gradual categorisation of body condition could help reduce the change in the odds of too thin mink with sub-period. The change with sub-period in the odds of too thin mink was not apparent when the prevalence of too thin mink was transformed into a measurement score for the winter period. This may be caused by the increased variation in the data caused by the non-linear transformation where the score for this measurement is reduced dramatically for a prevalence of up to approximately 10% and less thereafter (Møller *et al* 2015). This highlights the importance of an accurate assessment of the prevalence of too thin mink, as even small changes may have a large impact on the score.

As hypothesised, the odds of fur-chewing increased with sub-period with higher odds in sub-periods 2 and 3 compared to sub-period 1. This change affected the WelFur scores at all aggregated levels. However, the actual effect on the scores was relatively low at the higher levels. At principle level ('Appropriate behaviour'), only the score based on the prevalence from sub-period 2 was lower than sub-period 1, and the estimated difference was only one point. The assigned weights increase the diminishing effect of the aggregations, as the measurement score for 'Fur-chewing' in the winter period and its related scores at the higher levels are weighed less than the scores they are aggregated with (Møller *et al* 2015). There are implications for whether the measurement score for 'Fur-chewing' in the winter period and its related scores at the higher levels have the highest or lowest score in the aggregations, thus, the effect on calculated scores differs.

Contrary to what was hypothesised, there was no significant change in the odds of stereotypic mink in the winter period, but it should also be noted that the *P*-values were close to the border of significance, indicating that it could be worth investigating further. However, one reason could be that stereotypic behaviour was observed before feeding or, if this was not possible, as the last observation of the day. As 'the last observation of the day' can vary considerably, this has now been changed to 'from 1.5 hours before sunset' in order to increase the reliability of the assessment. Furthermore, a correction factor to adjust for difference in the prevalence of stereotypic behaviour 'before feeding'

and 'before sunset' is under development. Thus, the results regarding stereotypic behaviour in this study are, therefore, not necessarily comparable to how the assessment is carried out today. Hansen *et al* (2009) found that the prevalence of stereotypic behaviour observed during postponed feeding increased during January and February in a study where the mink were fed restrictively from December and the weight gradually decreased until the end of February. The eight farms included in the present study may have used different management strategies when slimming the mink, for example, the length of the slimming period and the degree of weight loss, which may have affected the minks' feeding motivation and, hence, the development and display of stereotypic behaviour. Studies comparing restrictive feeding to *ad libitum* feeding in the winter period have found that restrictive feeding increases the prevalence of stereotypic behaviour (Houbak & Møller 2000; Damgaard *et al* 2004). To our knowledge, there are no studies on how restrictive feeding is managed in practice and how differences in management affect the development and display of stereotypic behaviour.

When including the actual values for all variable measurements in the winter period, the score for 'Cage enrichment' in the winter period increased with sub-period. The measurement score for 'Cage enrichment' is aggregated with the measurement scores 'Fur-chewing' and 'Stereotypic behaviour' into the criteria score 'Expression of other behaviours'. However, the score for the criterion 'Expression of other behaviours' still decreased with sub-period. This is probably due to the aggregations where the lowest score is the starting point, and a low score can only be partly compensated for by higher scores in another area. In this case, the score for 'Cage enrichment' was higher than the scores for 'Fur-chewing' and 'Stereotypic behaviour', thus, the score for 'Cage enrichments' has the least impact on the aggregated score. The decreasing score for 'Fur-chewing' with sub-period, therefore, has a larger impact than the increasing score for 'Cage enrichment.' However, the changes in the principle score, 'Appropriate behaviour', were no longer apparent. This indicates that variation in other variable measurements masks some of the observed changes.

Growth period

The odds of mink with injuries in sub-period 4 were increased compared to sub-period 1, which supports our hypothesis. However, this change was not apparent when transformed into a measurement score for the growth period. The reason for this may be that this calculation also takes the severity of the injury into account; hence, an increased prevalence of the total number of injuries may not lead to an increased score for injuries if the severity is lower.

As hypothesised, the odds of cages with signs of diarrhoea increased with sub-period and were higher in sub-period 4 than in the first three sub-periods. The variation increased when the prevalence of cages with diarrhoea was transformed into a measurement score for the growth period due to the non-linear transformation where the score of welfare for this measurement is reduced dramatically for a prevalence of up to approximately 8% and less thereafter (Møller *et al* 2015). However, the WelFur scores at all aggregated levels were still affected by the change in the odds of cages with signs of diarrhoea, but the effect was relatively low at the higher levels. At principle level ('Good health'), the estimated score based on the prevalence from sub-period 4 was only decreased by 3 to 5 points compared to sub-periods 1, 2 and 3, while the scores on measurement level were decreased by 38 to 41 points. The assigned weights increase the diminishing effect of the aggregations as the measurement score for 'Diarrhoea' is not given the highest weight in the aggregation into the criteria 'Absence of disease' (Møller *et al* 2015). It differs whether the measurement score for 'Diarrhoea' in the growth period and its related scores at the higher levels have the highest or lowest value in the aggregations, thus, the effect on calculated scores differs.

The odds of exploratory mink increased with sub-period, thereby supporting the hypothesis. The prevalence of exploratory mink is aggregated with the prevalence of fearful and aggressive/undecided mink into the measurement score for 'Temperament test'. The increased odds of exploratory mink (and thereby decreased odds of fearful or undecided mink, as there were no aggressive mink) affected the WelFur scores at all levels but with a decreasing effect at the higher levels. At principle level ('Appropriate behaviour'), the score based on the prevalence from sub-periods 2, 3 and 4 was only increased 3 to 4 points compared to sub-period 1, compared to an increase of 12 to 16 points at measurement level. The assigned weights reduce the diminishing effect of the aggregations, as the weight given to the measurement score for 'Temperament test' in the growth period and its related scores at the higher levels is relatively high, thereby helping maintain the effect of the changes throughout the aggregations (Møller *et al* 2015). It differs whether the measurement score for 'Temperament test' in the growth period and its related scores at the higher levels have the highest or lowest value in the aggregations, thus, the effect on calculated scores differs. Due to the repeated testing of the mink throughout the period, there is a risk that the mink habituated to the test situation, thus increasing the prevalence of exploratory mink. Malmkvist and Hansen (2002) suggested that the reduced fear response of mink they found when testing juvenile mink weekly during the growth season in different fear tests was due to habituation to the test situations, even though the tests were not identical. In the present study, the majority of the tested mink were juveniles. However, due to the repeated testing it is not possible to separate the effect of habituation and the possible effect of maturation of the juvenile mink. This difference should be investigated further, as only an effect of the latter is a reliability issue in the WelFur system.

When including the actual values for all variable measurements in the growth period, the changes found in the WelFur scores, when only varying the measurements that were expected to change, were still present. This means that the changes in these measurements are not masked by variation in other measurements. The measurement 'Nest-box material and bedding/nesting material' was also found to change with

https://doi.org/10.7120/09627286.28.3.365 Published online by Cambridge University Press

sub-period with a higher score in sub-period 4 compared to sub-period 1. Management may have caused this measurement score to change as mink access to straw is often increased by the end of the growth period. This change also affected the aggregated criteria score for 'Thermal comfort' but not the principle score for 'Good housing'. The assigned weights increase the diminishing effect of the aggregations as the criteria score for 'Thermal comfort' is given the lowest weight in the aggregation into the principle score for 'Good housing' (Møller *et al* 2015). It matters whether the criteria score for 'Thermal comfort' has the highest or lowest value in the aggregation into the principle score for 'Good housing', thus, the effect on calculated scores differs.

Changes in overall evaluation

There were changes in the prevalence of welfare problems with date of assessment that could potentially lead to changes in the overall welfare categorisation. Some of the changes were positive in regards to animal welfare (decreased odds of too thin mink and increased odds of exploratory mink) and others were negative (increased odds of cages with diarrhoea, mink with injuries and fur-chewing). However, the resulting changes in the principle scores were limited. Later date of assessment in the winter period only led to a small change in the principle 'Appropriate behaviour' with a reduction of only one point in sub-period 2 compared to sub-period 1. Also, this change was not apparent when all variable measurements were included with their actual values. Later date of assessment in the growth period led to a decrease in the principle score for 'Good health' where the score in sub-period 4 was 3 to 5 points lower compared to the other periods and an increase in the principle score for 'Appropriate behaviour' where the score in sub-periods 2, 3 and 4 were 2 to 4 points higher than in sub-period 1. Except for one farm in the growth period, the overall categorisation of the farms did not change with date of assessment in the winter or growth periods. This shows that the assessment protocol is fairly robust to changes with date of assessment within the defined time windows. Each measurement taken on the farm in each period uncovers one aspect of the welfare, but one measurement alone should not explain the evaluation of the welfare at farm level. However, changes in the evaluation of individual measurements taken on the farm in each period should, to some extent, affect the evaluation at the higher levels as also shown for several measurements in this study. The aggregations control how much each score affects the evaluation of welfare at the higher levels. Higher scores can only partly compensate for a lower score, which means that changes in lower scores will have a larger effect than changes in higher scores. This also means that the effect of changes in the evaluation of individual measurement is affected by the results of other measurements. The compensation depends on the assigned weights; hence, changes in scores with a high weight will have a larger effect than changes in scores with a lower weight. Finally, as the number of measurements per criteria and number of criteria per principle vary, the more scores that are aggregated, the lower the effect of individual scores on the evaluation of welfare at the higher levels.

As discussed by Henriksen (2015) for changes in welfare in the nursing period, the overall categorisation of the farms could have changed if the principle scores were closer to the threshold between two categories. The changes in the principle scores in this study were smaller than the estimated changes with date of assessment for the principles 'Good feeding' and 'Good housing' in the nursing period, which were above 25 points (Henriksen & Møller 2015). In order to avoid changes with assessment date in the nursing period affecting the overall evaluation of welfare, Henriksen and Møller (2015) suggested stratifying the assessments between the assessment periods into the beginning, the middle and the end of each assessment period. However, this would require that change in welfare with assessment date in the three assessment periods largely followed the same pattern. Based on the results in this study, the changes in welfare with date of assessment in the winter period have only a limited effect on the overall evaluation. Also, the changes in the growth period point in opposite directions, as the changes with date of assessment for the principle 'Good health' are negative as regards animal welfare and the changes for the principle 'Appropriate behaviour' positive. The suggested stratification of the assessments does, therefore, not seem an adequate solution to avoiding changes with date of assessment at measurement level affecting the overall categorisation. An alternative solution could be to shorten the time window for each assessment period in order to limit the variation. This would, however, seriously challenge the ongoing assessment of the approximately 3,000 mink farms in Europe that was initiated in 2017. A more feasible way could, therefore, be to develop a correction factor for the few measurements that can be expected to change within each assessment period. Finally, one could simply accept the documented changes with date of assessment and appreciate the fact that the effects on the principle scores and overall category are limited as they have turned out to be here.

Animal welfare implications

In a seasonally synchronised production system, such as mink production, it is important to include every season in the assessment, since different seasons may be associated with specific animal groups and welfare risk factors (Møller *et al* 2003). But, also, date of assessment within each season has to be considered, as there may be rapid changes within the season which may affect the outcome of the assessment as highlighted by this study. As the results of the WelFur assessment are directly available to the farmers to be used as part of their daily management, the possible changes in welfare with date of assessment, as found in our study, may be important for the farmers, even if it does not affect the overall evaluation. Knowing the effect of date of assessment gives the farmer the possibility of interpreting the results as regards to the risk factors that were present or absent on the day of the assessment. This must also be considered when assessing the welfare in other production systems with seasonal breeders as, for example, sheep production, or in production systems with

non-seasonal breeders kept in seasonal breeding systems, as in the case of dairy cows in grazing systems and spring calving. In production systems where all the animals on the farm are the same age on the day of the assessment as, for example, with laying hens, this is also important.

Conclusion

There was a change in the odds of welfare problems with date of assessment for all of the expected measurement variables, apart from stereotypic behaviour in the winter period. Also, the change in the odds of too thin mink was the reverse of what would have been expected, as the odds decreased. Some changes with date of assessment were positive as regards animal welfare and others negative. However, these changes had only a limited effect on WelFur scores at the higher levels. Apart from one farm in the growth period, the final categorisation of the farms did not change with date of assessment. However, if the scores were closer to the threshold between two categories, this could have been the case for more farms. This has to be taken into account in the WelFur-Mink assessment in order to ensure a valid assessment independent of assessment date. It seems that each assessment period must be addressed individually since the changes in welfare with assessment date differ in direction and magnitude between each of the three assessment periods.

Acknowledgements

We thank all the farmers for kindly letting us onto their farms. We also thank Elna Mortensen and Maria Blæsbjerg-Obitsø, both Kopenhagen Fur, for their help with data collection and valuable comments. Steen Henrik Møller was responsible for study design and Bente Krogh Hansen for data collection. Anna Feldberg Marsbøll was responsible for data analysis which was completed with help from Leslie Foldager, Aarhus University. All authors contributed to the manuscript. Fur Europe and Aarhus University financed this project.

References

Bates D, Maechler M, Bolker B and Walker S 2015 Fitting linear mixed-effects models using lme4. *Journal of Statistical Software 67*: 1-48. https://doi.org/10.18637/jss.v067.i01

Botreau R, Veissier I, Butterworth A, Bracke MBM and Keeling, LJ 2007 Definition of criteria for overall assessment of animal welfare. *Animal Welfare 16*: 225-228

Damgaard BM, Hansen SW, Børsting CF and Møller SH 2004 Effects of different feeding strategies during the winter period on behaviour and performance in mink females (*Mustela vison*). *Applied Animal Behaviour Science 89*: 163-180. https://doi.org/10.1016/j.applanim.2004.04.010

Hansen M 1985 Diseases and hygiene. In: Jørgensen G (ed) *Mink Production* pp 261-340. Scientifur. http://www.ifasanet.org

Hansen SW 1996 Selection for behavioural traits in farm mink. *Applied Animal Behaviour Science 49*: 137-148. https://doi.org/10.1016/0168-1591(96)01045-3

Hansen SW, Clausen TN and Sandbøl P 2009 Winter feeding to reduce stereotypies and increase reproduction. In: Sandbøl P (ed) *Annual Report 2008* pp 7-18, Danish Fur Breeders Research Center: Holstebro, Demark

Hansen SW, Møller SH and Damgaard BM 2014 Bite marks in mink: induced experimentally and as reflection of aggressive encounters between mink. *Applied Animal Behaviour Science 158*: 76-85. https://doi.org/10.1016/j.applanim.2014.06.008

Henriksen BIF 2015 *Assessment of animal welfare in mink farms - based on the WelFur-Mink Protocol*. PhD Thesis, Aarhus University, Denmark

Henriksen BIF and Møller SH 2015 The reliability of welfare assessment according to the WelFur-protocol in the nursing period of mink (*Neovison vison*) is challenged by increasing welfare problems prior to weaning. *Animal Welfare 24*: 193-201. https://doi.org/10.7120/09627286.24.2.193

Hothorn T, Bretz F and Westfall P 2008 Simultaneous inference in general parametric models. *Biometrical Journal 50*: 346-363. https://doi.org/10.1002/bimj.200810425

Houbak B and Møller SH 2000 Activity and stereotypic behaviour in mink dams fed *ad libitum* or restricted during the winter. In: Murphy BD and Lohi U (eds) *Proceedings of the VIIth International Scientific Congress in Fur Animal Production* pp 146-150. 13-15 September 2000, Kastoria, Greece

Hunter B 1996 Digestive system of mink. In: Hunter DB and Lemieux N (eds) *Mink… Biology, Health and Disease* pp 14.1-14.20. University of Guelph: Ontario, Canada

Malmkvist J and Hansen SW 2001 The welfare of farmed mink (*Mustela vison*) in relation to behavioural selection: A review. *Animal Welfare 10*: 41-52

Malmkvist J and Hansen SW 2002 Generalisation of fear in farm mink, *Mustela vison*, genetically selected for behaviour towards humans. *Animal Behaviour 64*: 487-501. https://doi.org/10.1006/anbe.2002.3058

Malmkvist J, Palme R, Svendsen PM and Hansen SW 2013 Additional foraging elements reduce abnormal behaviour – fur-chewing and stereotypic behaviour – in farmed mink (*Neovison vison*). *Applied Animal Behaviour Science 149*: 77-86. https://doi.org/10.1016/j.applanim.2013.10.001

Møller SH, Hansen SW, Malmkvist J, Vinke CM, Lidfors L, Gaborit M and Botreau R 2015 *WelFur Welfare assessment protocol for mink*. European Fur Breeders' Association: Brussels, Belgium

Møller SH, Hansen SW and Sørensen JT 2003 Assessing animal welfare in a strictly synchronous production system: The Mink Case. *Animal Welfare 12*: 699-703

Mononen J, Møller SH, Hansen SW, Hovland AL, Koistinen T, Lidfors L, Malmkvist J, Vinke CM and Ahola L 2012 The development of on-farm welfare assessment protocols for foxes and mink: the WelFur project. *Animal Welfare 21*: 363-371. https://doi.org/10.7120/09627286.21.3.363

R Core Team 2017 *R: A language and environment for statistical computing*. R Foundation for Statistical Computing: Vienna, Austria. https://www.R-project.org